DOCUMENT RESUME

ED 397 119 TM 025 254

AUTHOR Halkitis, Perry N.; And Others
TITLE Estimating Testing Time: The Effects of Item Characteristics on Response Latency.
PUB DATE Apr 96
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12, 1996).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Computer Assisted Testing; Difficulty Level; *Estimation (Mathematics); Licensing Examinations (Professions); *Prediction; Property Appraisal; *Responses; Test Construction; *Test Items; *Timed Tests; Time Factors (Learning)
IDENTIFIERS *Item Discrimination (Tests); *Speededness (Tests); Word Counts

ABSTRACT

The relationship between test item characteristics and testing time was studied for a computer-administered licensing examination. One objective of the study was to develop a model to predict testing time on the basis of known item characteristics. Response latencies (i.e., the amount of time taken by examinees to read, review, and answer items) were obtained from 197 individuals taking a national licensing examination for real estate appraisers for the first time. Response latencies were measured by the computer during the test taking process. Results of the study, which parallel results from paper-and-pencil tests, indicate that item response time was determined by three item characteristics: (1) item difficulty; (2) item discrimination; and (3) word count. These variables accounted for about half the variance in the response time. Item position, however, seems to have an inverse effect on response time, indicating that less time is required as one progresses through the examination. This may be due to the effect of practice as the examinee gains more experience with computer testing or with test speededness. Results provide a temporary model that test developers can use in estimating the amount of time that should be allotted to computer-administered examinations. (Contains two tables and seven references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY
PERRY N. HALKITIS

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

ED 397 119

Estimating Testing Time: The Effects of Item Characteristics on Response Latency

Perry N. Halkitis, PhD
Gay Men's Health Crisis, Inc.
129 West 20th Street
New York, NY 10011-0029

J. Patrick Jones, PhD
Jitendra Pradhan, MA
Professional Examination Service
475 Riverside Drive
New York, NY 10115-0089

Paper presented at the annual meeting of the American Educational Research Association in New York, New York. For information or copies of this paper, please contact the first author.

TM025254

BEST COPY AVAILABLE

Introduction

In the last decade, computer-administered tests of ability and achievement have gained in popularity. Proponents of computer-administered tests cite the following advantages: greater test administration standardization, improved test security, enhanced item display capability, and reduced testing time (Bunderson, Inouye, & Olson, 1989). The use of computers for test administration activities also permits direct measurement of response latencies [e.g., item stem and response exposure times, response selection time, etc.] (Bunderson et al, 1989). While considerable research has been conducted on the development and scoring of computer-administered tests (cf., Kiely, Zara, & Weiss, 1986; Millman, 1977), less attention has been devoted to test administration issues such as testing time.

Previous research on testing time for computer-administered examinations has focused on differential effects of computer administration versus paper-and-pencil administration (cf., Bugbee & Bernt, 1992; Olsen, Maynes, Slawson, & Ho, 1986; Wise & Plake, 1989) and the use of item response theory for item analysis and test scoring (Wainer, 1983). In a review of the literature on computer-administered examinations, Wise and Plake suggested that examinees may require less time to complete multiple-choice items administered by computer, as compared to paper-and-pencil tests. Olsen et al. also reported a significant reduction in testing time for elementary school students who completed a computer-administered educational achievement test. In contrast to these findings, Bugbee and Bernt observed that computer-administered test takers required significantly more time to finish certification tests in financial services than did

paper-and-pencil test takers.

Based on previous research, there is no consistent trend regarding testing time differences between computer-administered tests and paper-and-pencil tests, and properties of test items that may effect testing time (e.g., difficulty, discrimination, length, position in test, etc.) for computer-administered tests have not been systematically investigated. Accurate projections of the amount of testing time required by examinees are necessary to ensure that unintended effects due to response speed do not compromise score interpretations from computer-administered tests. The purpose of this study was to examine the relationship between test item characteristics and testing time for a computer-administered licensing examination. One objective of this investigation was to develop a model to predict testing time on the basis of known item characteristics.

METHOD

Response latencies (i.e., the amount of time taken by examinees to read, review, and answer items) were obtained from individuals testing for the first time with one level of a national licensing examination for real estate appraisers. The examination consists of 100 four-option, multiple-choice items, and candidates are allowed 15 minutes to gain familiarity with computer-administered testing procedures and 2 hours and 45 minutes to complete the test.

Candidates were administered the licensing examination on microcomputers at Drake Authorized Testing Centers (DATC's) throughout the United States. In this system, candidates enter responses using either the computer keyboard or a pointing device (i.e., a "mouse"). The examination was administered as a fixed test form, but

the sequence of items was randomly presenfed to test takers.

Response latencies are measured directly by the DATC microcomputers, and these data are stored with candidate item responses. At the conclusion of testing, candidate data files were transferred to the investigators via modem.

For each examinee, a data file containing item responses, response latencies, item position on the rest, and item word length were compiled. Item difficulty estimates (i.e., percentage of candidates selecting the keyed response) and item discrimination indexes (i.e., point biserial correlation coefficient) were computed for each test item. On this examination consisting of 100 items mean item difficulty was 0.78 and mean discrimination was 0.23. On the DATC system response latency was clocked from the second that the item has completed plotting on the computer until the candidate presses "next" to move to the next item on the examination. If an examinee failed to answer and item and the returned to it later, total time on the item was accumulated across the exam administration. Word count for each item was determined using the this facility in a word processing package.

Since response latencies are typically positively skewed, a logarithmic transformation was applied to item response latencies before any data analyses were completed.

It was anticipated that a linear relationship would exist between item difficulty, item discrimination, item word length, and response latency. To describe this linear relationship, a multiple regression analysis was performed to predict response latency on the basis of item difficulty, item discrimination, and item word length.

To determine the impact of item position on response latency, average response

latencies were computed by item position on the test. Rank-order correlations between average response latency and item position were calculated, and a discrete graph constructed to examine the relationship between these variables.

RESULTS

The variables of interest in the study were the dependent variable, response latency (res), and the independent variables, item difficulty (p), item discrimination (r), and word length of the actual item (wl). Descriptive statistics for these items are provided in Table 1. These statistics were based on administration of the examination to 197 U.S. candidates tested in 1995.

Table1
Descriptive Statistics of Variables of Interest

VARIABLE	N	MEAN	ST DEV	MIN	MAX
RESP TIME (res)	100	77.15	75.52	16.80	455.86
DIFFICULTY (p)	100	0.78	0.17	0.21	0.99
DISCRIM (r)	100	0.23	0.13	-0.07	0.49
WORD LNG (wl)	100	44.76	21.47	17.00	106.00

Since response latencies tended to be positively skewed, the logarithm of res was calculated prior to any analyses. The resulting mean of logarithmic response time (logres) was 1.77 (s=.29).

Initial analyses included the computation of a correlation matrix for all variable in the analyses. This correlation matrix was constructed to investigate multicollinearity

among independent variables ib the regression model. While the independent variables were not found to be significantly related to each other, all three independent variables were significantly related to the outcome variable logres (logarithm of the response latency). The correlation between word length (wl) and logres was 0.52; the correlation between item difficulty (p) and logres was -0.43; and the correlation between logres and item discrimination (r) 0.31. Each correlation was significant at $p<.01$. The correlation coefficients suggest that as item length and item discrimination increase so does response latency. As the item becomes more difficult, so does the response time.

Regression analyses of logres on wl, p, and r was undertaken using a forward approach. Regression analysis results in a significant linear relationship ($F=32.23$, $p<.0001$) with the model accounting for 50.18% of the variance in logres. The parameter estimates, the partial R-squares, and their associated significance level are reported in Table 2.

Table 2
Variables Entered in Regression Analysis

VARIABLE	PARAMETER ESTIMATE	PARTIAL R-SQUARE	F (p)
wl	.00644	27.2%	43.41 (p<.0001)
p	-.00687	16.2%	30.52 (p<.0001)
r	.57977	6.8%	13.14 (p<.0001)
INTERCEPT	1.88858		262.14 (p<.0005)

In addition to the above analyses, the relationship between item position on the examination and response time was examined. A Spearman rank-order correlation analysis was conducted between the rank order of mean response all items in *nth* position and the sequential position of items in the test. A statistically significant negative relationship between mean response time and item position was observed ($r=-.38$, $p<.0001$) suggesting that as completes more items, the response time decreases.

DISCUSSION

Insufficient testing time represents a potential source of invalidity. The determination of testing time for credentialing examinations is a decision that will have a significant impact on test validity , testing efficiency, and resource allocation. The use of computers for test administration activities provides a unique opportunity to measure response latency and systematically examine factors that affect testing time.

Results of the study indicate that item response time on a computer administered examination is determined in part by three item characteristics--item

difficulty, item discrimination, and word count, with these variables accounting for about half the variance in response time. These results parallel those found on paper-and-pencil examinations. Item position, itself, seems to have an inverse effect on response time, indicating that less time is required as one progresses through the examination. This finding may be due to the practice effects as the examinee gains more experience with computer testing or with test speededness.

The results from this study provide a preliminary model that can be used by test developers in estimating the time that should be allotted to computer-administered examinations. Prior to the advent of computer based testing, these time estimates have been based on traditional paper-and-pencil exam results where response latencies are not readily available. Given an item's world length and its psychometric characteristics (item difficulty and discrimination), an initial estimate for total testing time can be generated using a regression model.

References

Bugbee, A.C., & Bernt, F.M. (1992). Times up! The effects of time constraints and mode of administration on test performance and perception. *Journal on Computing in Education. 25*(2),243-253.

Bunderson, C.V., Inouye, D.K., & Olson, J.B. (1989). The four generations of computerized educational measurement. In R.L. Linn (Ed.), *Educational measurement* (3rd ed.), New York: American Council on Education-MacMillan, 367-407.

Kiely, G.D., Zara, A.R., & Weiss, D.J. (1986), *Equivalence of computer and paper-and-pencil Armed Services Vocational Aptitude Battery Tests* (Research Report No. AFHRL-TP-86-13). Brooks Air Force Base, TX; Air Force Human Resources Laboratory.

Millman, J. (1977, April). *Creating domain referenced tests by computer.* Paper presented at the meeting of the American Educational Research Association, New York.

Olsen, J.B., Maynes, D.M., Ho, K. & Slawson, D.A. (1986). *The development and pilot testing of a comprehensive assessment system, phase I.* Provo, UT. Waterford Testing Center.

Wainer, H (1983). On item response theory and computerized adaptive tests. *Journal of College Admissions*, 27, 9-16.

Wise, S.L., & Plake, B.S. (1989). Research on the effects of administering tests via computers. *Educational Measurement.* Issues and Practice, *8*(3), 5-10.